EVALUATION OF THE 'ELECTRONIC FEEDBACK' MARKING ASSISTANT AND ANALYSIS OF A NOVEL COLLUSION DETECTION FACILITY

Phil Denton

Evaluation of the 'Electronic Feedback' Marking Assistant and Analysis of a Novel Collusion Detection Facility

Philip Denton
Faculty Learning Development Manager
Faculty of Science
Byrom St
Liverpool JMU
Liverpool
L3 3AF
UK

Tel: 0151 231 2035
Fax: 0151 231 2170
Email: p.denton@livjm.ac.uk

Abstract

Electronic Feedback is an MS Excel/Word marking assistant that has been developed by the author over the past 4 years. After the tutor inputs appropriate data, the system is able to generate and email feedback reports to students. Such reports can include so-called standard comments, selected from a bank of statements that are anticipated to be required regularly during marking. The program also calculates allocation statistics, being the proportion of students in a class who required each standard comment. This article reports the results of the first formal evaluation of the software, a questionnaire returned by 22 users of Electronic Feedback 8. Over a range of disciplines and class sizes, academics agreed that they were able to return more feedback, of higher quality, and in a shorter space of time, when using Electronic Feedback. The allocation statistics were considered to be of use to tutors, presumably because they can help the assessor identify precisely those aspects of an assessment that caused particular difficulty. In response to this survey, the program has been completely redesigned so that it is less intimidating to those unfamiliar with Excel. Version 9 now includes a main menu and an improved interface for allocating standard comments to students. Tutors have welcomed these modifications, the number of users rising to 44 over the last academic year. The updated software also incorporates a novel collusion detection facility, capable of finding pairs of students who required similar feedback comments. Using this approach, the scripts of identified students would then be re-examined, to confirm whether the likeness was coincidental or suspicious. The results of applying this procedure to 11 assessments at JMU are discussed. It is concluded that the system can detect instances of plagiarism in free text, even hand-written work, although further work is required to establish the limitations of this approach.

Introduction

To date, CAA has tended to concentrate on objective testing; indeed the two terms are often used interchangeably. A more complete description of CAA would include:

- objective tests,
- data collection tasks, including database interrogation and web-surfing,
- electronic submission of work, including web-supported threaded discussions,
- automated analysis of students' free-text work, including plagiarism detection and paperless marking systems,
- marking assistants, including packages that can be used to email feedback to students on free text assignments.

Ramsden reports that students attach great value to high quality feedback, considering it to be of greater importance than clear explanations and the stimulation of interest within the classroom (Ramsden, 1993). Of course, the educational benefit of even the most well crafted feedback is lost if it is returned too late (Gibbs, 1993). It is unsurprising, therefore, that studies in this area indicate that an absence of appropriate feedback is an important contributory cause of student failure (Entwistle, 1989).

Despite their potential value, marking assistants are not widely used in UK HE. Commenting on the results of a CAA Centre survey, Stephens remarks that the vast majority of marking is still done by hand (Stephens, 2001). Typically, CAA is preferred for formative and diagnostic testing (Martin, 2001), while summative, 'high stakes' assessments do not benefit from electronic assistance. Stephens goes on to suggest a reason for this, citing the, "... absence of any front-running software in this regard, compared to the high profile of Question Mark." Most of the existing report-writing assistants, for example, are designed for the US High School market.

Electronic Feedback is an MS Office program that Stephens describes as a 'home grown' marking tool, having been developed by the author at Liverpool JMU over the last 4 years. The operation of Version 5 and Version 8 (Denton, 2001) have been described previously. Although based on MS Excel and Word, the system is fully functionalised by extensive MS Visual Basic for Applications (VBA) programming. A principle feature of the program are the standard comments, a bank of feedback statements that are anticipated to be required repeatedly during marking. Tutors may compose these in advance of marking, or build up a crib-sheet of comments as marking proceeds. Each remark has an associated reference number, allowing for speedy allocation to individual students. Standard comments can be arranged as a simple list (normal mode) or organised under discrete sub-headings (criterion mode). In normal mode, the software presents information, allocation statistics, indicating the % of class members that required each comment.

The purpose of this article is to report on the results of the first formal survey of those UK academics supplied with Version 8. The paper will also summarise recent advancements in the program, Version 9, including a novel collusion detection facility. Details of this approach have been published previously (Denton, 2002), but the results of applying the program 'in the field' are reported here for the first time. The article ends with a consideration of future work, including a preview of Version 10, due for release in July 2003.

Evaluation of Version 8

Method

A questionnaire was emailed to staff at 59 HE institutions in 2002. All staff had been able to use Electronic Feedback 8 software for up to 10 months. Respondents were selected because they had either attended a presentation of the program, took part in a training session, or had requested the software via email. Typically, training sessions and presentations were organised in collaboration with institutional staff development centres. Generally, these were open to all university teaching staff. Email requests for the software were usually prompted by these presentations.

Staff were given 3 weeks to complete the survey and 142 replies were received. In addition, the author has had regular, informal conversations with users of the software. Certain questions required tutors to pick their favoured response from a Likert scale: 1 = 'strongly agree', 2 = 'agree', 3 = 'neutral', 4 = 'disagree', 5 = 'strongly disagree' (Barnett, 1991).

Results and Discussion

Some responses to the questionnaire are reported in Tables 1 and 2. A glance at the first Table suggests that the take up-rate of the software is poor, 22 users out of 142 staff. As will be discussed later, this is not as poor as it might first appear. Although the software was originally developed to assist in the marking of laboratory reports, the survey indicates that it has now been used across a range of disciplines. There was insufficient raw data to investigate whether there were any variations in how the software was used from subject to subject.

Given the initial investment of time required to familiarise oneself with any computerised system, it might be expected that the software would be favoured by tutors of large groups (>40). In fact, there is no evidence to support this notion. The average class size for users and non-users alike is around 55. The approach of one advocate of the program suggests a reason for this. They use the software with particularly small groups (<10) and cite 3 reasons:

- The facility to enter a personal comment to individual class members.
- They prefer to type, rather than hand-write feedback (Interestingly, all of the 142 surveyed academics use MS Word at least twice a week).
- The emailing facility, which enables the return of feedback, even in the absence of formal contact time.

Gender	Users	Non-users
Male	12	73
Female	10	47
Subject Specialism		
Arts and Humanities	5	23
Computing, Maths and Engineering	3	36
Physical and Life Sciences	6	21
Medicine and Nursing	2	11
Other and Unknown	6	29

Table 1. Gender and subject specialisms of users and non-users of Electronic Feedback 8

A fuller understanding of why tutors prefer to use the program is supplied by the answers to the questions reported in Table 2. Pleasingly, it was found that tutors agreed that they were able to return more detailed and higher quality feedback, in a shorter period of time, when using the software. Tutors also agreed that the allocation statistics were useful. By directing the tutor's attention to those aspects of an assessment that caused particular difficulty, the software could be used to inform future teaching, learning and assessment strategies. Users can also elect to incorporate the allocation statistics into the feedback reports generated by the program. Tutors were less sure, however, as to whether or not this information was of use to students.

	Question	Mean	SD	no.
1	Using the Electronic Feedback software, I have been able to return...			
	a) more feedback	2.1	1.1	18
	b) higher quality feedback	1.6	0.8	18
	c) feedback more quickly	2.1	1.2	18
2	The allocation statistics are useful for tutors	1.9	0.7	17
3	The allocation statistics are useful for students	2.8	1.0	17
4	I have found Electronic Feedback 8 easy to use	1.8	0.6	18
5	My colleagues could use the software without formal training	2.9	1.2	18

Table 2. Averaged Likert scale responses to questions set to 22 users of the Electronic Feedback 8 software. 1 = 'Strongly agree' to 5 = 'strongly disagree'.

The take-up rate of the program, 22 out of 142, represents 15% of the surveyed group. However, this rises to 22% when only those staff who attended a 'hands-on' training workshop are considered. Staff who just attended a talk may not have 'actively' volunteered to be present, but would have received the software anyway. Accordingly, their take-up rate is only 12%. The survey identified the principle factor that determined whether staff chose to use the software. On average, users of Electronic Feedback 8 interact with MS Excel almost one a week. For non-users, this increases to nearer once every 7 weeks. Non-users agreed that 'a lack of time' was the principle reason why they had not taken-up the software. This is understandable, if academics felt that they had to familiarise themselves with MS Excel, before even interacting with Electronic Feedback itself. This is perhaps also the reason why even users of the program were only neutral when asked to comment as to whether their colleagues could use the program without formal training, Table 1. There are other reasons, however, one academic deciding not to use the software because they were unable to obtain student names and email addresses in an electronic format.

Brief Description of Version 9 and Analysis of a Novel Collusion Detection Facility

Version 9 was released in July 2002. A detailed description will not be given here as the basics of the method are as described elsewhere (Denton, 2001). Only novel features will be considered in this paper.

Method

Electronic Feedback 9 consists of three files:

- Guide9.xls, an interactive guide to the software,
- Feedback9.xls, into which marks, comments and student details are inputted,
- Fb9.doc, an MS Word document that formats and emails the feedback reports.

These 3 files must be saved within the same folder. Whereas Version 8 had very much the appearance on a MS Excel file, Version 9 has increased functionality and includes a main menu. The sheets that compose this new workbook are accessed via this menu, and not by clicking on the sheet name tabs, as in Version 8 and any 'normal' MS Excel file. Figure 1 shows the main menu from Version 10, which has the same principle function as the main menu of the preceding version and will be released in July 2003, after Version 9 expires.

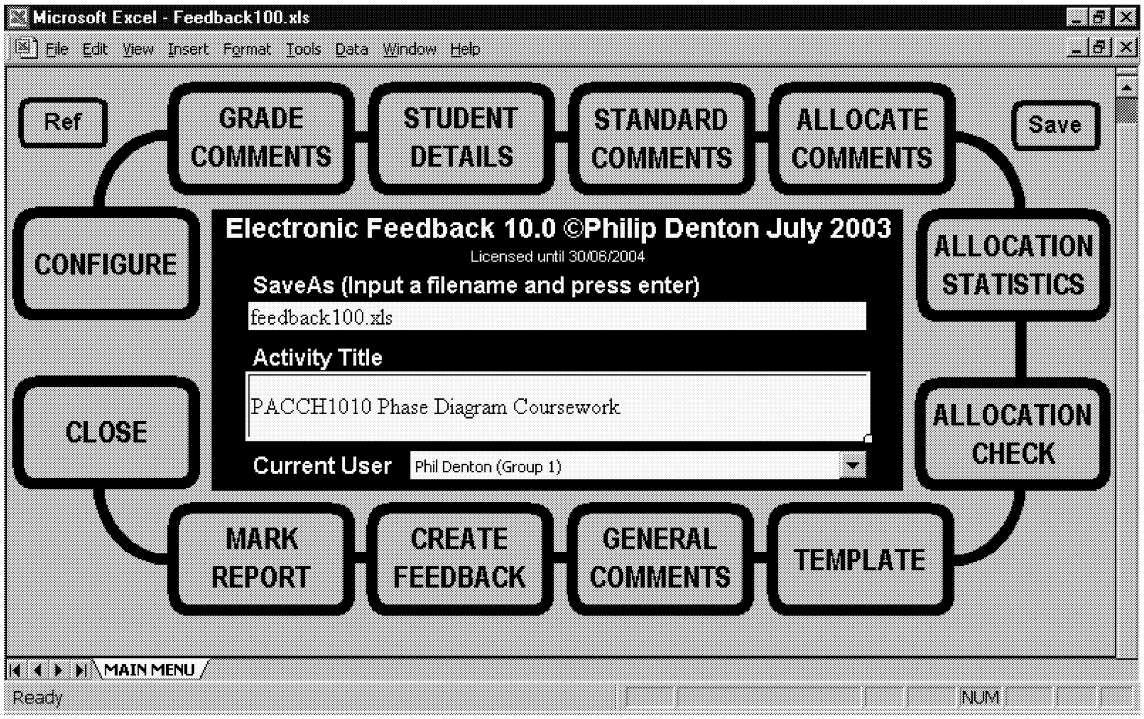

Figure 1. Main menu of Electronic Feedback Version 10

Among the new features introduced in Version 9:

- Extensive use of drop down menus to facilitate rapid data entry.
- An Import/Export feature that allows files of comments and student details to be readily saved and retrieved.
- The facility to enter up to 101 separate grade comments. This now allows borderline remarks, such as, “First/2(i) Class” to be accommodated.
- A ‘salutation’ field on the student details sheet, allowing the student’s preferred moniker to be inputted. The special character, ‘~’, may then be incorporated when composing comments and indicates that the student’s salutation is to be inserted at that point. This allows for the personalisation of even general comments, to the entire class.
- A new Allocate Comments sheet. Here, the desired student is selected, and the awarded mark, personal comment, and standard comments are inputted. As before, standard comments are selected by entering the associated reference number. In Version 9, however, tutors can also select the comment itself from a drop-down menu. This has the advantage that the tutor need not recall the associated reference numbers for each comment.
- A collusion detection facility, allocation check, that works on the principle that students’ requiring similar standard comments may well have a significant likeness in their work that merits further examination.

In a draft copy of Version 9, not distributed to tutors, the allocation check feature was able to identify groups of students who had been allocated identical standard comments. The ability of this system to detect collusion was tested with a series of free text chemistry assessments, marked by the author.

Results and Discussion

The results of applying the allocation check procedure to 11 assessments at JMU are shown in Table 3. Five tests resulted in the detection of students with identical standard comments. 5 of the 6 groups identified by assessment 4 had worked together during the practical element of their laboratory report. Upon inspection, it was concluded that similarities in the work amounted to acceptable collaboration, given the fact that students were reporting the same results. The scripts of the remaining group were not suspicious.

Assessment	Type	No. of students	Total no. of standard comments	Average no. of comments allocated to each student	No. of groups of students with identical standard comments
1	Laboratory report	18	15	8.4	0
2	Laboratory report	19	14	7.7	0
3	Laboratory report	19	20	7.9	0
4	Laboratory report	20	10	3.2	6
5	Structured Questions	23	21	9.8	2
6	Structured Questions	24	27	11.0	0
7	Structured Questions	26	28	14.6	0
8	Laboratory report	47	27	10.3	0
9	Laboratory report	48	22	13.5	4
10	Structured Questions[a]	58	53	15	5
11	Structured Questions	56	25	5.9	2

[a]Criterion Mode. For each student, an appropriate comment was chosen for each of the 15 questions set.

Table 3. Results of applying the allocation check facility to assignments submitted by Level 1 and 2 chemistry students, Liverpool JMU, 2001-2003. Normal mode, except were indicated.

In assessment 5, feedback given to students on a series of structured worksheet questions was analysed. Two pairs of students had identical feedback, although it was considered that the work of one pair had an unintentional likeness. The other couple, however, had genuine similarities in their scripts that had been missed because the work had been marked a few days apart. It was considered that these similarities were on the borderline between acceptable and unacceptable collusion. Consequently, a warning message was returned in the personal comment to each of the students concerned, using Electronic Feedback.

Assessment 9 considered feedback returned on a practical exercise. 4 groups of students with identical comments were detected. Two groups had similarities in the computer-generated aspects of their assignments, but not the hand-written sections. The similarities of the third group's work were considered to be coincidental. The fourth group's assignments were not checked as the scripts had been returned to the students.

In assessment 10, 5 groups had been awarded identical comments, in response to a series of structured mathematical exercises. 3 of these were considered to be coincidental, after inspection of the original work. The fourth group of students had sufficient similarities in their work to merit a written warning. In the fifth case, however, it was considered that the submitted work constituted unacceptable collusion. The two students concerned were required to meet with the author to discuss the matter. It was concluded that the plagiarism was not intentional, but had arisen because the group had worked together and had shared answers, believing that this was permitted. Given that these were Level 1 students, and that this was an assignment set early in Semester 1, it was considered appropriate to give a verbal warning regarding the students' future conduct. The information supplied to students with this assessment next year will be revised to ensure that students understand what constitutes acceptable group working.

In assessment 11, 2 pairs of students were found to have identical comments, but their work was not considered to have a suspicious likeness. This failure to find plagiarism highlights a deficiency in the draft software. Anecdotal evidence suggests that colluding students purposefully introduce inconsistencies in their scripts so that they cannot be accused of cheating. In addition, a student that attempts to transcribe another's work may unintentionally introduce errors that necessitate different feedback to the source script. To address this problem, the final version of the software features customisable detection criteria. Thus, in normal mode, tutors can search for groups of students who have at least X standard comments in common, and no more than Y standard comments that are different. X and Y are integers.

The search for possible collusion in assessment 11 was repeated, this time using the final edition of Version 9 to search for student pairs with at least 5 standard comments in common, and no more than 1 comment that was different. These parameters were chosen by trial and error, after a series of searches, until the number of identified pairs was within reasonable limits. 9 couples matched the detection criteria, and the results of the search are displayed in a tabular format. A right-mouse click on any entry in this table activates a message box that gives further information on the selected pair. The message box shown in Figure 2 relates to two students whose written work showed clear signs of collusion, upon inspection.

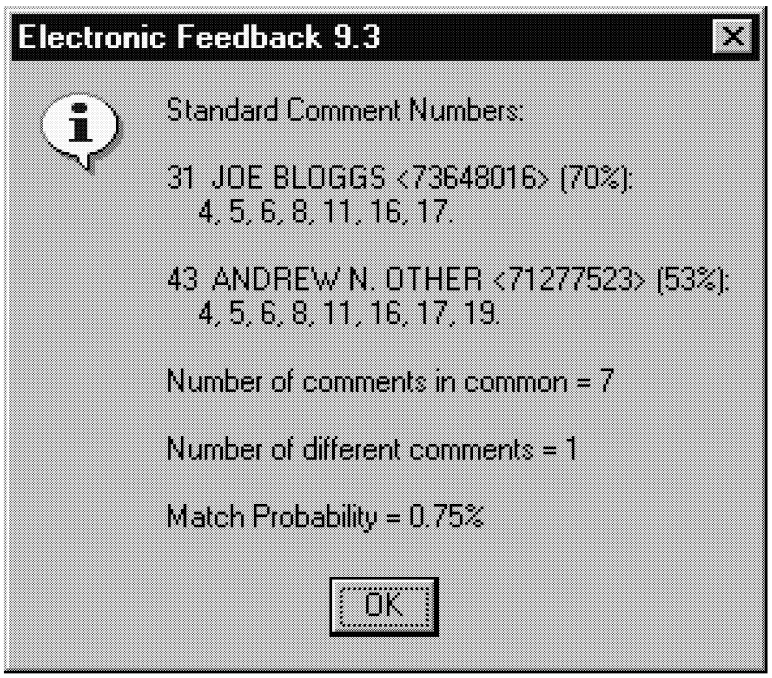

Figure 2. Example message box generated by allocation check

In Figure 2, '31' and '43' relate to the position of the students on the class list. Student registration numbers are also shown, along with the awarded % mark, in brackets. It is clear that these two students have seven standard comments in common, with reference numbers, 4, 5, 6, 8, 11, 16, and 17. In addition, standard comment number 19 was awarded to one of the students. The match probability is calculated by multiplying together the allocation statistics for each of the matching comments. Thus, if this pair of students shared 7 comments, but each comment was awarded to only four-fifths of the class, the match probability would be $100*(0.8)^7$ = 21.0%. Effectively, the match probability is a measure of the likelihood that the similarities in two students' scripts are coincidental. A very low figure, such as that in Figure 2, suggests to the marker that this pair of scripts merits inspection.

Overall Conclusions

The results of the questionnaire suggest that the Electronic Feedback concept is educationally sound, but that Version 8 required modification to make it less intimidating to users unfamiliar with MS Excel. In accord with this finding, nearly 70% of the surveyed academics expressed an interest in revised versions of the program. Already, initial work suggests that the amendments incorporated into Version 9 have enhanced the attractiveness of the program. An emailed question in May 2003, asked only if tutors were using Version 9. This received 44 replies in the affirmative, although a further 42 colleagues said they intended to investigate the software.

The theory underpinning the collusion detection facility is sound, given that this approach has successfully detected instances of plagiarism. The ability to customise detection criteria allows collusion to be detected that would otherwise be missed. Future work will investigate the limitations of this approach. Other research projects may consider:

- The value to students of returning to them the allocation statistics, showing the frequency with which individual feedback comments were used by the marker.
- The implications of using the software in light of the recent Special Educational Needs and Disability Act (SENDA). Use of the software could overcome difficulties associated with the legibility of staff handwriting by visually impaired students. Moreover, the program could be of use to those staff who have difficulty handwriting feedback remarks.
- Effects on marking consistency resulting from the use of a common set of feedback comments in those assessments were there is more than one marker.
- Examples of how use of the program has lead to changes in course design.

It is hoped that Version 10 of the program will be available in July 2003. As before, the software will made freely available to academic staff and will expire in one year's time. In this way, the author has only to deal with enquires about one version of the software at any one time. The principle changes that will be incorporated into the next release are an improved capability to deal with more that one assessor, and the facility to input standard comments in both criterion and normal modes. Long-term, it would be hoped that the software could be developed into a wholly web-based system that could be integrated into existing institutional databases.

Additional Information

Any colleague wishing to obtain a copy of the software should email the author.

Acknowledgement

This work has been supported by a JMU Teaching Fellowship.

References

Barnett, V. (1991) *Sample Survey Principles and Methods 2nd Edition* Oxford: Oxford University Press.

Brown G, Bull J and Pendlebury M (1997) *Assessing Student Learning in Higher Education* Routledge, London, 7.

Burnstein J, Leacock C and Swartz R (2001) Automated Evaluation of Essays and Sort Answers *Proceedings of the 5th International Computer Assisted Assessment Conference*, 41-54.

Denton, P. (2001) Generating and e-Mailing Feedback to Students Using MS Office *Proceedings of the 5th International Computer Aided Assessment Conference,* 157-173.

Denton, P. (2001) Generating Coursework Feedback for Large Groups of Students Using MS Excel and MS Word *U. Chem. Ed.* 5, 1-8.

Denton, P. (2002) Detection of Plagiarism in Students' Written Work Using a Novel Electronic Marking Assistant, *Proceedings of the International Conference on ICTs in Education*, 445-449.

Entwistle, N.J., Hounsell, D.J., Macaulay, C., Situnayake, G. and Tait, H. (1989) *The performance of Electrical Engineers in Scottish Higher Education*, Report to Scottish Education Department, Centre for Research on Learning and Instruction, Department of Education, University of Edinburgh.

Gibbs, G. and Habeshaw, T. (1993) *Preparing to Teach: An Introduction to Effective Teaching in Higher Education* Bristol: Technical and Educational Services Bristol.

Martin, N. (2001) *Web-based assessment.* Testing Times. **1** (4) 1-7.

Ramsden, P. (1993) *Learning to Teach in Higher Education* London: Routledge.

Stephens, D.P., Sargent, G. and Brew, B. (2001) *Comparison of assessed work marking software: implication for the ideal Integrated Marking Tool (IMT)* Proceedings of the. 5th International Computer Assisted Assessment Conference, pp 471-485.